THEORY-GROUNDED EVALUATION OF HUMAN-LIKE FALLACY PATTERNS IN LLM REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

We study logical reasoning in language models by asking whether their errors follow established human fallacy patterns. Using the Erotetic Theory of Reasoning (ETR) and its open-source implementation, PyETR, we programmatically generate 383 formally specified reasoning problems and evaluate 38 models. For each response, we judge logical correctness and, when incorrect, whether it matches an ETR-predicted fallacy. Two results stand out: (i) as a capability proxy (Chatbot Arena Elo) increases, a larger share of a model's incorrect answers are ETR-predicted fallacies ($\rho = 0.360, p = 0.0265$), while overall correctness on this dataset shows no correlation with capability; (ii) reversing premise order significantly reduces fallacy production for many models, mirroring human order effects. Methodologically, PyETR provides an open-source pipeline for unbounded, synthetic, contamination-resistant reasoning tests linked to a cognitive theory, enabling analyses that focus on error composition rather than error rate.

1 INTRODUCTION

Language models increasingly solve complex tasks Wei et al. (2023); Kojima et al. (2022). We ask a simple question: when they err on controlled reasoning problems, do their errors align with human fallacies? Human reasoning shows systematic, repeatable fallacies Tversky & Kahneman (1974); Kahneman & Tversky (1982); Evans (1994); Walsh & Johnson-Laird (2004); Johnson-Laird (2006). These are not random mistakes; they follow predictable patterns across tasks. Understanding whether LLM errors share these patterns is useful both scientifically and for deployment in settings that require reliable reasoning Jacobs (2021); Bommasani et al. (2022).

To rigorously investigate this question, we leverage the Erotetic Theory of Reasoning (ETR) Koralus (2022); Mascarenhas & Koralus (2017a); Koralus & Mascarenhas (2013), a formal cognitive theory that precisely predicts human reasoning patterns across multiple domains. ETR posits that human reasoning operates by maintaining disjunctive alternatives and filtering these alternatives when taking on new information, a process that can systematically lead to error. ETR provides formal specifications of both how and when humans will make specific reasoning errors, allowing us to generate arbitrary reasoning problems with predictable failure patterns.

ETR has an open-source implementation, PyETR Koralus et al. (2025). Using PyETR, we mechanically generate reasoning problems predicted to elicit specific fallacies. We evaluate 38 models on a fixed corpus of 383 such problems.

Our experimental investigation evaluates 38 language models ranging from smaller models (e.g., `Mistral 7B Instruct v0.1`) to larger systems (e.g., `GPT-4.5`, `Claude 3.7`), treating Chatbot Arena Elo as one capability proxy Chiang et al. (2024). We observe that as this proxy increases, a larger share of logically incorrect answers align with ETR-predicted human fallacies ($\rho = 0.360, p = 0.0265$). We do not claim causality and restrict interpretation to this evaluation setting.

While language models often improve on many benchmarks Wei et al. (2023); Bubeck et al. (2023), our results indicate that within our task domain, the proportion of errors matching ETR-predicted fallacies increases with Chatbot Arena Elo as a capability proxy. These findings are consistent with overlap between LLM error patterns and human fallacy patterns and do not require commitments about underlying cognitive mechanisms.

We check the robustness of our result by using metrics other than Chatbot Arena Elo where available for a subset of the 38 models. We found significant Spearman correlation with estimates of training compute Epoch AI (2025) ($\rho = 0.489, p = 0.0334$) and significant exponential fit for mean score on the HELM capabilities benchmark Liang et al. (2023) ($r = 0.796, p = 0.0103$).

Contributions We report a statistically significant correlation between Chatbot Arena Elo as a capability proxy and reasoning errors predicted by ETR, an empirically grounded formal theory of human reasoning. We also provide the first substantial application of PyETR — an open-source implementation of ETR — to generate unbounded, synthetic, contamination-resistant reasoning tasks linking model errors to theory-predicted fallacies.

The remainder of the paper is organised as follows. Section 2 introduces ETR and PyETR. Section 3 details methodology. Section 4 presents results. Section 5 discusses implications.

Related Work Substantial prior literature in psychology and cognitive science documents irrationalities in human judgement and decision-making Tversky & Kahneman (1974); Kahneman & Tversky (1982); Evans (1994); Kahneman et al. (1990). In particular, the psychology of deductive inference reveals human propositional reasoning to be vulnerable to logically-irrelevant linguistic effects, like co-references between premises Walsh & Johnson-Laird (2004). Mental model theory provides an explanatory paradigm for these results Johnson-Laird & Byrne (1991), though prior work on the Erotetic Theory of Reasoning matches and has replicated mental model predictions on deductive inference tasks Mascarenhas & Koralus (2015; 2017b).

In the language model literature, prior studies find that while modern LLMs largely succeed at syllogistic inference Clark et al. (2021); Eisape et al. (2024); Bertolazzi et al. (2024), they exhibit human-like reasoning failures, including distractability Shi et al. (2023), content effects Lampinen et al. (2024), and order effects Eisape et al. (2024); Saparov & He (2023). Closest to our work, the first application of the Erotetic Theory to LLMs found the predictive utility of the theory to increase for larger model sizes Koralus & Wang-Maścianica (2023), though this study was limited to OpenAI's GPT model family Brown et al. (2020); OpenAI (2024; 2022) and did not make use of PyETR. Our implementation yields an endlessly regenerative and content-agnostic test of the Erotetic Theory for LLMs in a manner more resilient to data contamination.

2 The Erotetic Theory of Reason

Modern accounts of human reasoning highlight a duality: impressive competence and predictable errors. The Erotetic Theory of Reasoning (ETR) explains both with one idea: we reason by managing questions and candidate answers ("alternatives") and filtering them when new information arrives. Filtering is efficient but can drop relevant alternatives too early, producing characteristic fallacies. ETR formalises this process and defines when conclusions stabilise against follow-up questions ("erotetic equilibrium").

At a high level (full formalism in Appendix A), ETR has three ingredients: (1) maintain disjunctive alternatives as candidate answers to an implicit question, (2) filter by best match with incoming information (risking premature elimination), and (3) recover by asking questions (raising structured alternatives) that reintroduce necessary information. This predicts a range of empirical reasoning data across propositional and first-order logic, probability, and decision-making. Morover, it is mathematically proved in Koralus (2022) that ETR converges to normative standards of rationality in the limit, as the recovery step introduces enough alternatives to stabilise further reasoning steps.

Let us demonstrate ETR with a compact example (Example 49 in Koralus (2022)).

Premise 1	Either there is an ace in Mary's hand and some other player has a king, or else there is a queen in John's hand and some other player has a jack.
Premise 2	Sally has a king.
Question	Does it follow that Mary has an ace?

The correct logical answer is "not necessarily", yet many respondents endorse "Mary has an ace" (60% in data following Mascarenhas & Koralus (2017a)).

To model the problem with logic, we would first choose a representation of the basic predicates, in this case $\mathsf{Ace}(x)$ for the predicate that (the card) x is an ace, and $\mathsf{Has}(y, z)$ for the predicate that (player) y holds (a card) z. The same applies when using ETR, where instead of logical formulas we have the central concept of *view* (see Appendix A) for the full mathematical formalism). It will suffice for now to use the (still precise) shorthand notation to express the premises and conclusion as ETR views:[1]

P1	$\{\mathsf{Ace}(a^*)\mathsf{Has}(\mathsf{Mary}, a)\mathsf{King}(b^*)\mathsf{Has}(c, b), \mathsf{Queen}(d)\mathsf{Has}(\mathsf{John}, d)\mathsf{Jack}(e)\mathsf{Has}(f, e)\}$
P2	$\{\mathsf{King}(g^*)\mathsf{Has}(\mathsf{Sally}, g)\}$
Q	$\{\mathsf{Ace}(h^*)\mathsf{Has}(\mathsf{Mary}, h)\}$

Each view is given as a disjunctive set of alternatives, separated by a comma. Each alternative is to be considered as a conjunctive set of atomic propositions, where for convenience we omit the set-brackets and commas and simply juxtapose elements. Many readers will recognise these as *disjunctive normal forms*, but in ETR there is no inherent ordering to the atoms in a conjunction or alternatives in the disjunction. The individual lowercase letters can be understood are as existentially quantified variables, with the scope of the quantification encompassing the entire disjunction (‘prenex form’). A view may also carry a supposition, written as a superscript (see Table 1 for examples). Note that since any formula of first-order logic can be written in prenex disjunctive normal form, ETR is as expressive as full first-order logic.

PyETR, an open-source software package that serves as a calculator for ETR, can parse views from a similar text format.

```
from pyetr import View
p1 = View.from_str("∃a ∃b ∃c ∃d ∃e ∃f {Ace(a*)Has(Mary(),a)Has(c,b
    )King(b*),Has(John(),d)Has(f,e)Jack(e)Queen(d)}")
p2 = View.from_str("∃g {King(g*)Has(Sally(),g)}")
cc = View.from_str("∃h {Ace(h*)Has(Mary(),h)}")
```

Note the requirement to explicitly declare existentially quantified names (the letter `E` may be used in place of ∃). ETR posits some default procedures for ordinary reasoning, built out of some basic operations. The workhorse among the basic operations is Update, which update a current view with an incoming one. We can check in PyETR that ETR does predict a fallacy here, because Update discards the seemingly less relevant alternative where John has a queen.

```
>>> v = p1.update(p2); print(v)
∃g ∃l ∃m {Ace(l*)Has(Mary(),l)Has(Sally(),g)Has(m,g)King(g*)}
>>> v.query(cc)
∃h {Ace(h*)Has(Mary(),h)}
```

The actual default procedure of ETR involves a few more steps which become relevant in more complex problems. The full version is implemented in PyETR.

```
>>> from pyetr.inference import default_procedure_does_it_follow
>>> default_procedure_does_it_follow([p1,p2],cc)
True
```

3 METHODOLOGY

In our experiment, a *reasoning problem* is given by a list of views (the *premises*), and the problem is to answer the question “what if anything follows?”. Our method generates reasoning problems where the ETR-predicted answer is a logical fallacy.

[1]An additional aspect not present in ordinary logic is *issue structure*. This is notated with an asterisk on some occurrences of terms, indicative of being ‘at issue’ for the containing atomic proposition. The issue structure has no logical content, but plays a role in guiding ETR inferences. Determination of issue structure from cues in natural language or context is outside the scope of ETR, but a simple heuristic in this example is that the repetition of ‘Ace’ and ‘King’ suggests those concepts are at issue.

3.1 Data processing

Table 1: Original bank of reasoning problems.

Problem	Symbolic representation	Natural language example
Modus ponens	$\{R(x)\}^{\{Q(x)\}}$ $\{Q(x)\}$ $\therefore \{R(x)\}$	If it rains, it's wet. It rains. Therefore, it's wet.
Modus tollens	$\{R(x)\}^{\{Q(x)\}}$ $\{\overline{R(x)}\}$ $\therefore \{\overline{Q(x)}\}$	If the switch is on, then the lamp is lit. The lamp is not lit. Therefore, the switch is not on.
Quantified modus ponens	$\forall x\{R(x)\}^{\{Q(x)\}}$ $\forall x\{Q(x)\}^{\{P(x)\}}$ $\therefore \forall x\{R(x)\}^{\{P(x)\}}$	All mammals have lungs. All dogs are mammals. Therefore, all dogs have lungs.
Disjunction fallacy	$\{Q(x)R(x), S(x)T(x)\}$ $\{Q(x)\}$ $\therefore \{R(x)\}$	She is quiet and clever, or tall and athletic. She is quiet. Therefore, she is clever.

We source our original bank of reasoning problems from examples presented in the *Reason and Inquiry* text (Koralus, 2022). These include basic "template" reasoning problems like modus ponens, modus tollens, and the disjunction fallacy as originally presented in Walsh & Johnson-Laird (2004). Table 1 presents the complete list of original problems. Not all reasoning problems take the form of valid syllogistic inferences; that is, not all conclusions from Erotetic Theory inference deductively follow from the premises.

To avoid complications in the mapping to natural language, we restrict to views with only *monadic* (one-place) predicates. The corresponding monadic first-order logic is known to be less expressive than full first-order logic, but still allows for a rich set of reasoning problems. Our method could easily be extended to include polyadic predicates.

3.2 Generative pipeline

We define a collection of mutation functions that give slight modifications to ETR view objects. These functions serve to expand our bank of reasoning problems by modifying existing problems by single premises. Table 2 presents the complete list of possible mutation rules.

Table 2: Logical Form Mutations

Mutation Type	Description
Predicate Addition	Introduces a new predicate symbol to the language.
Constant Addition	Introduces a new constant symbol to the formal system.
Variable Addition	Adds a new variable to the set of arbitrary objects.
Constant-to-Variable Substitution	Replaces constants with $\forall$ or $\exists$ quantified variables.
Conjunctive Atom Insertion	Conjoins new atomic formulas to existing states.
Disjunctive State Addition	Disjunctively creates new states with atoms.
Atom Negation	Applies or removes negation from atomic formulas.

Generating a new problem proceeds by iterative building a list of views (the premises in the problem). A new view is selected at random from the original problem bank and subjected to a random number of random mutation rules, it is added to the list if ETR predicts the problem admits a non-trivial answer. New views are added until all of the following stopping conditions are met. (1) The problem is the right size, meaning that summing the number of atoms in each view gives between 4 and 11. If 11 is exceeded, backtracking is used. (2) The ETR-predicted conclusion contains a single alternative (no disjunctions). (3) The ETR-predicted conclusion is a logical fallacy. To mitigate concerns about dataset-specific effects, we intentionally generated problems across diverse domains and structures. The PyETR-based generation ensures our problems test reasoning capacities rather than memorization of training examples.

Dataset curation We initially generated 400 problems. Pre-analysis integrity checks identified 17 items whose fallacy status did not meet the final specification, so we excluded them prior to all analyses, yielding 383 problems. Given the modest reduction in sample size and the cost of re-running all model evaluations, we report results on this vetted 383-set; the generation pipeline supports regenerating larger sets for future work.

3.3 MAPPING TO NATURAL LANGUAGE

To evaluate logical reasoning in language models in a more naturalistic context, we developed a systematic approach to convert formal logical statements into natural language prompts. We created 12 themes for natural language framings of logical problems, such as a researcher figuring out the properties of a novel element or a newly discovered creature. For each theme, we established consistent mappings between logical elements and thematic equivalents. The thematic conversion process serves multiple purposes: it ensures our results are robust against content effects Lampinen et al. (2024); it enables us to test whether the reasoning patterns are consistent in diverse contexts; and it mitigates potential data contamination, a significant problem for LLM evaluation Zhou et al. (2023); Deng et al. (2024); Li & Flanigan (2024); Singh et al. (2024) by reformulating problems in novel scenarios unlikely to appear in training data.

Consider the following example of a disjunction fallacy problem in its formal representation: $\mathbf{P1} : \{Q(x)R(y), S(x)T(y)\}, \mathbf{P2} : \{Q(x)\}$. In the "alchemy" theme, logical predicates (e.g., $A(x)$, $B(y)$) were mapped to thematic attributes (e.g., "is transmuting", "is time-bending"), while logical variables (e.g., x, y) were mapped to imaginary thematic entities (e.g., "cosmic dust", "vital mercury"). Using our alchemy theme, this was transformed into:

```
I'm an alchemist studying mysterious substances in my laboratory.
   I need to understand their properties through logical
   reasoning. Here's what I've discovered:

Cosmic dust is transmuting and vital mercury is time-bending, or
   cosmic dust is immortality-granting and vital mercury is
   spirit-affecting.
Also, we know that cosmic dust is immortality-granting.

What if anything follows? I do not have an intended answer in mind
   , and it is possible that nothing follows. Please be succinct
   and precise.
```

This is typical of the prompts sent to the LLM. It includes a thematic preamble and a natural language prompt to reason in an open-ended fashion with: "What if anything follows?" Each problem maintained the same prompt structure: an introduction establishing the thematic context, followed by the premises expressed in natural language, and concluding with a standardised question asking what follows from the premises. The prompt explicitly stated that nothing might follow, to avoid biasing models toward making unwarranted inferences Saparov & He (2023).

3.4 DATA COLLECTION

Model Selection and Testing Conditions We evaluated 38 language models of varying sophistication, including several specifically designed as reasoning models. Selection informally prioritized quantity, popularity of usage, and representation of a diverse range of scores. All models were accessed via OpenRouter and tested using LLMharness. Chatbot Arena scores were obtained from `https://huggingface.co/spaces/lmarena-ai/`

chatbot-arena-leaderboard/resolve/main/elo_results_20250423.pkl. Scores for the HELM Capabilities benchmark for 9 models were obtained from https://crfm.stanford.edu/helm/capabilities/v1.10.0/#/leaderboard Liang et al. (2023) and estimates for training compute for 19 models were obtained from https://epoch.ai/data/ai-models Epoch AI (2025) (both URLs accessed 30 July 2025).

To ensure a fair comparison across architectural differences, all reasoning models were allocated 2,400 thinking tokens during evaluation. This standardization was justified given our use of the Chatbot Arena Elo score as our primary capability metric, which already accounts for performance differences regardless of underlying mechanisms Chiang et al. (2024).

Response Processing and Filtering Models provided conclusions in natural language, which we converted to PyETR format for evaluation. To ensure fair assessment of reasoning (not formatting), GPT-4.1-mini served as our translation layer without access to the original premises. Spot-checking confirmed translation fidelity.

Each model was evaluated on 383 reasoning tasks generated by PyETR. During analysis, we encountered occasional parse errors when processing model responses. To maintain data quality while preserving statistical power, models with parse error rates exceeding 20% across the test set were excluded from analysis, which resulted in the exclusion of one model (google__gemini-2.5-pro-preview-03-25). For the remaining models, individual responses that produced parse errors were excluded from analysis on a case-by-case basis rather than discarding all responses from that model.

Testing Framework The full set of logical problems were framed as evals, using Eleuther's Language Model Evaluation Harness Gao et al. (2024), which handles evaling and metric collection against a wide range of models. OpenRouter OpenRouter (2025) was used for all models, in order to have a unified API. This consumed less than $1,000 of compute resources. Each model was given a token limit of 3,000 output tokens.

Key Measures We collected the following measures to evaluate both the logical correctness of model responses and their alignment with human-like reasoning patterns, enabling our analysis of how these patterns correlate with model sophistication. A model's answer was considered **'logically correct'** if it was a logical consequence of the premises. This was tested using PySMT Cimatti et al. (2017) to check whether the negation of the conclusion was inconsistent with the premises. LLM responses were classified as **'ETR-predicted'** if ETR predicts an endorsement of that conclusion according to PyETR's implementation of default_procedure_does_it_follow.

Statistical Tests We selected Pearson correlation to capture linear relationships between variables while also reporting Spearman's rank correlation to account for potentially non-linear monotonic relationships without assuming normality in the distribution of fallacy rates. For our premise-order intervention analysis, we employed two-proportion z-tests to rigorously evaluate whether reversing premise order produced statistically significant changes in fallacy rates. The z-test was specifically chosen because it allows us to determine if the observed differences in proportions (original vs. reversed premise order) represent genuine effects rather than random variation, providing a reliable measure of the significance of order effects across different model capabilities. This statistical approach enables us to quantify the extent to which models exhibit the same premise-order sensitivity documented in human reasoning studies.

4 Experimental Results

Our analysis reveals several notable patterns in how language models exhibit human-like reasoning fallacies. We report the results of Pearson correlation tests, as well as Spearman tests for the detection of non-parametric monotonic relationships. We consider $p \leq 0.05$ a suitable boundary for statistical significance.

Proportion of Errors that are Fallacies We operationalise a "human-like fallacy" (or simply "fallacy") as a response that is both ETR-predicted and logically incorrect. This definition captures instances where models make the same systematic reasoning errors that ETR predicts humans would make. Formally, for a given model m and reasoning problem p, we define:

$$\text{HumanLikeFallacy}(m, p) = \begin{cases} 1 & \text{if ETR-predicted}(m, p) \wedge \neg\text{LogicallyCorrect}(m, p) \\ 0 & \text{otherwise} \end{cases}$$

Our metric of interest is the *fallacy rate*, defined as the proportion of human-like fallacies to total logically incorrect answers for each model:

$$\text{FallacyRate}(m) = \frac{\sum_{p \in P} \text{HumanLikeFallacy}(m, p)}{\sum_{p \in P} \neg\text{LogicallyCorrect}(m, p)}$$

where P is the set of all reasoning problems in our evaluation set. This measure represents the degree to which a model's errors align with predictable human reasoning patterns rather than other forms of logical error. **We find as model strength increases, so does the proportion of logically incorrect answers that are fallacies** (Figure 1).

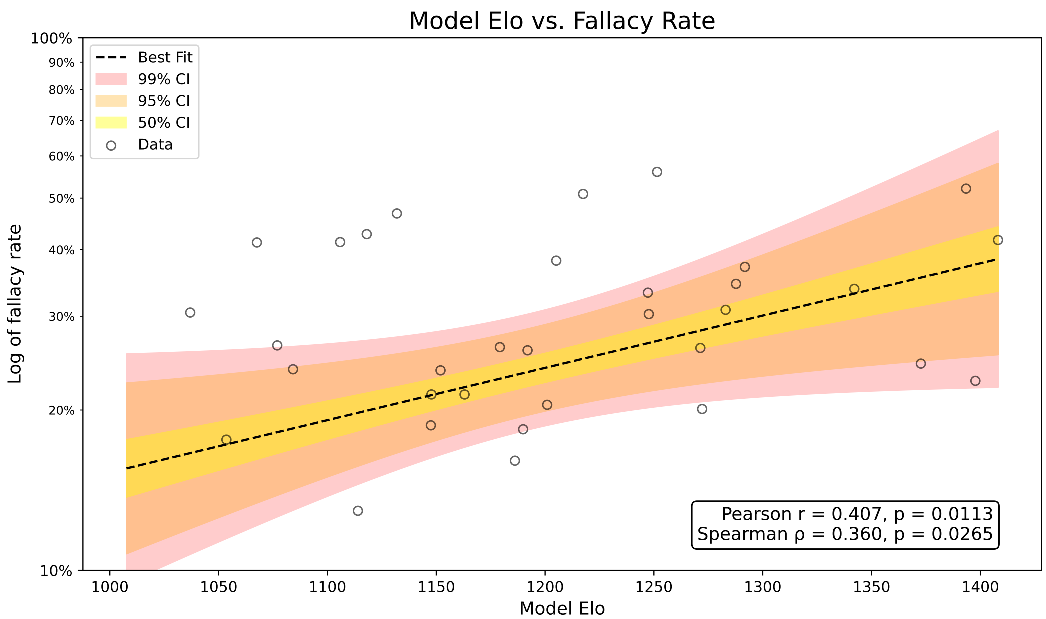

Figure 1: Over 38 models, a direct linear fit produces a significant moderate positive correlation ($r = 0.364, p = \mathbf{0.0247}$), an exponential fit (log y-axis) displays a more confident Pearson correlation ($r = 0.407, p = \mathbf{0.0113}$), as displayed above. As Spearman is ordinal, we have the same significant monotonic relationship in both cases ($\rho = 0.360, p = \mathbf{0.0265}$).

Overall Correctness and other measures Surprisingly, model strength as measured by LLM Elo score does not appear to have any correlation ($r = 0.004, p = 0.981$), ($\rho = -0.04, p = 0.777$) with the model's ability to produce correct answers on our dataset: see Table 3. Furthermore, exploratory analysis did not reveal any other statistically significant evidence of model Elo correlating or enjoying a monotonic ordinal relationship with absolute proportion of fallacies produced, whether exactly predicted by ETR or taking logical equivalences.

Table 3: Percentage of model responses which are logically correct

Mean	σ	**min**	$\mathcal{Q}(0.25)$	$\mathcal{Q}(0.5)$	$\mathcal{Q}(0.75)$	**max**
40.6%	16.7%	18.6%	31.0%	37.5%	44.8%	91.7%

Intervention study: Reversing Premise Order While in classical logic the order in which premises are presented has no effect on normatively correct logical conclusions, an *order-effect* is commonly observed in human reasoners Girotto et al. (1997), where presenting premises in different orders elicits logically different responses. In addition to being a proxy measure for the non-classicality of human reasoners, this non-commutativity effect often blocks fallacies. We observe strong evidence that LLMs are similar in this regard: presenting the same questions with reversed premise order significantly blocks fallacy production. Table 4 reports a significant fallacy-blocking effect from reversing premises on a majority of models.

Table 4: 38 models selected for testing, listed by Chatbot Arena Score (general). The right-most column reports the percentage of fallacies per-model blocked (answer becomes logically correct) by reversing premise order, along with the results of a two-proportion z-test for each model to gauge the significance of the blocking effect.

Provider	**Model**	**Elo**	**Fallacies blocked by reversal**
OpenAI	chatgpt-4o-latest	1408	46.15%, (z = 3.24, p = **1.18e-03**)
OpenAI	gpt-4.5-preview	1398	68.00%, (z = 3.03, p = **2.44e-03**)
Google	gemini-2.5-flash-preview	1393	37.27%, (z = 3.52, p = **4.27e-04**)
DeepSeek	deepseek-chat-v3-0324	1373	38.78%, (z = 2.28, p = **2.27e-02**)
Google	gemma-3-27b-it	1342	20.39%, (z = 1.77, p = 7.60e-02)
OpenAI	o1-mini	1304	50.00%, (z = 0.58, p = 5.63e-01)
Anthropic	claude-3.7-sonnet	1292	50.00%, (z = 4.23, p = **2.35e-05**)
xAI	grok-2-1212	1288	22.68%, (z = 1.91, p = 5.58e-02)
Anthropic	claude-3.5-sonnet	1283	65.08%, (z = 4.74, p = **2.09e-06**)
OpenAI	gpt-4o-mini-2024-07-18	1272	36.36%, (z = 2.00, p = **4.59e-02**)
Google	gemini-flash-1.5	1271	29.69%, (z = 1.98, p = **4.78e-02**)
Mistral	mistral-large-2407	1251	37.60%, (z = 3.85, p = **1.18e-04**)
Meta	llama-3.1-70b-instruct	1248	24.32%, (z = 1.74, p = **8.14e-02**)
Anthropic	claude-3-opus	1247	32.63%, (z = 2.79, p = **5.20e-03**)
Mistral	mistral-small-24b-instruct-2501	1217	34.53%, (z = 3.81, p = **1.40e-04**)
Microsoft	phi-4	1205	44.68%, (z = 3.91, p = **9.34e-05**)
Anthropic	claude-3-sonnet	1201	53.49%, (z = 3.04, p = **2.40e-03**)
Google	gemma-2-9b-it	1192	55.17%, (z = 2.55, p = **1.07e-02**)
Cohere	command-r-plus-04-2024	1190	37.25%, (z = 2.22, p = **2.64e-02**)
OpenAI	gpt-4-0314	1186	41.67%, (z = 2.07, p = **3.81e-02**)
Anthropic	claude-3-haiku	1179	56.67%, (z = 3.91, p = **9.20e-05**)
OpenAI	gpt-4	1163	54.05%, (z = 2.85, p = **4.43e-03**)
Meta	llama-3-8b-instruct	1152	23.81%, (z = 1.55, p = 1.21e-01)
Mistral	mistral-medium	1148	43.90%, (z = 2.38, p = **1.75e-02**)
Mistral	mixtral-8x22b-instruct	1148	42.86%, (z = 2.12, p = **3.43e-02**)
Anthropic	claude-2.0	1132	32.81%, (z = 3.40, p = **6.65e-04**)
Anthropic	claude-2.1	1118	54.79%, (z = 4.28, p = **1.91e-05**)
Mistral	mixtral-8x7b-instruct	1114	37.50%, (z = 1.49, p = 1.36e-01)
OpenAI	gpt-3.5-turbo-0125	1106	39.42%, (z = 3.61, p = **3.06e-04**)
Meta	llama-3.2-3b-instruct	1103	20.00%, (z = 0.77, p = 4.41e-01)
Nous Research	nous-hermes-2-mixtral-8x7b-dpo	1084	66.67%, (z = 2.91, p = **3.63e-03**)
DeepSeek	deepseek-chat	1077	35.00%, (z = 2.29, p = **2.19e-02**)
Mistral	mistral-7b-instruct-v0.2	1072	38.46%, (z = 1.11, p = 2.66e-01)
OpenAI	gpt-3.5-turbo-1106	1068	88.46%, (z = 4.36, p = **1.33e-05**)
Meta	llama-3.2-1b-instruct	1054	32.43%, (z = 1.60, p = 1.10e-01)
Microsoft	phi-3-mini-128k-instruct	1037	38.16%, (z = 2.88, p = **3.96e-03**)
AllenAI	olmo-7b-instruct	1015	45.45%, (z = 1.23, p = 2.18e-01)
Mistral	mistral-7b-instruct-v0.1	1008	66.67%, (z = 1.42, p = 1.55e-01)

5 Conclusion and Discussion

Interpretation of Key Findings Our results reveal a statistically significant correlational trend in error composition: as language models advance in capability (proxied by Chatbot Arena Elo), a larger share of logically incorrect answers align with human errors predicted by ETR. While stronger language models demonstrate improved capabilities across many benchmarks, we do not observe increased overall correctness in this dataset. Moreover, we demonstrate order-effects for LLM reasoning analogous to those observed in humans. This indicates overlap between LLM error patterns and established human fallacy patterns on these tasks.

Methodological Considerations and Limitations While traditional ablation studies might seem appropriate for investigating this phenomenon, they face significant limitations in the current landscape of language models: the diversity of modern architectures, training approaches such as distillation Gou et al. (2021); Gupta & Agrawal (2022); Xu et al. (2024), and reasoning capabilities with test-time scaling DeepSeek-AI (2025); El-Kishky et al. (2024); Xu et al. (2025) means that simple parameters like model size may no longer serve as reliable proxies for model capability. Our use of Chatbot Arena scores provides a performance-based metric that naturally incorporates a diverse range of models, reflecting real-world ability rather than architectural specifics, similar to Ruan et al. (2024).

Though the correlation strength ($\rho = 0.360$) may appear moderate, it is robust across diverse model architectures and training paradigms, suggesting a fundamental relationship rather than artifact. The statistical significance ($p = 0.0265$) exceeds conventional thresholds despite our conservative analytic approach.

While our results demonstrate correlation between model capability and human-like errors, we acknowledge limitations in attributing causality. Alternative explanations include the possibility that stronger models are increasingly trained on human-produced reasoning traces that themselves contain these fallacies. The absence of improved logical accuracy with model capability may reflect ceiling effects in our test set or fundamental limitations in current training paradigms.

Theoretical Implications These findings challenge the assumption that scaling alone leads to more normatively correct reasoning systems. Instead, within our evaluation setting we observe greater overlap between model error patterns and characteristic human fallacies. This does not by itself imply shared internal processes; it highlights an evaluation axis (error composition) complementary to overall accuracy.

Our results provide a foundation for several research directions. For benchmark development, our approach offers a systematic method for generating reasoning problems where human-like fallacies are predicted, creating evaluation sets that could complement existing reasoning benchmarks. For capabilities, the identified fallacy patterns could serve as targeted training examples to fortify models against human-like reasoning errors while maintaining their general capabilities. As a tool for assessing engineering choices, our methodology provides a quantitative framework for evaluating the effectiveness of various interventions aimed at improving model reasoning, from prompt engineering to architectural modifications.

Broader Implications This observed trade-off in error composition has implications for AI alignment. As models become more capable, anticipating and mitigating human-like reasoning errors becomes increasingly important, especially in high-stakes contexts requiring reliable reasoning, such as medical diagnosis, legal analysis, or decision support Jacobs (2021); Bommasani et al. (2022).

We see these results as opening new avenues for understanding both artificial and human reasoning. As language models continue to advance, this work provides a foundation for developing systems that combine human-like understanding with more robust reasoning capabilities: systems that can successfully reason as we do while avoiding the pitfalls that characterise human cognition.

Author use of AI tools During manuscript preparation, we used AI assistance for prose shortening and copy-editing, and as coding support for scaffolding evaluation scripts and refactoring utilities. All analysis code, evaluation protocols, and results were designed, implemented, and verified by the authors.

References

Leonardo Bertolazzi, Albert Gatt, and Raffaella Bernardi. A systematic analysis of large language models as soft reasoners: The case of syllogistic inferences. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 13882–13905, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.769. URL https://aclanthology.org/2024.emnlp-main.769/.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models, 2022. URL https://arxiv.org/abs/2108.07258.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. Sparks of Artificial General Intelligence: Early experiments with GPT-4, 2023. URL http://arxiv.org/abs/2303.12712.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. Chatbot arena: An open platform for evaluating LLMs by human preference. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=3MW8GKNyzI.

Alessandro Cimatti, Alberto Griggio, Roberto Sebastiani, and Marco Roveri. PySMT: A python library for SMT solving. https://github.com/pysmt/pysmt, 2017. Accessed: 2025-05-15.

Peter Clark, Oyvind Tafjord, and Kyle Richardson. Transformers as soft reasoners over language. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 3882–3890, 2021.

DeepSeek-AI. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, 2025. URL https://arxiv.org/abs/2501.12948.

Chunyuan Deng, Yilun Zhao, Xiangru Tang, Mark Gerstein, and Arman Cohan. Investigating data contamination in modern benchmarks for large language models. In Kevin Duh, Helena Gomez, and Steven Bethard (eds.), *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 8706–8719. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.naacl-long.482. URL https://aclanthology.org/2024.naacl-long.482/.

Tiwalayo Eisape, Michael Tessler, Ishita Dasgupta, Fei Sha, Sjoerd Steenkiste, and Tal Linzen. A systematic comparison of syllogistic reasoning in humans and language models. In Kevin Duh, Helena Gomez, and Steven Bethard (eds.), *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 8425–8444, Mexico City, Mexico, June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.naacl-long.466. URL https://aclanthology.org/2024.naacl-long.466/.

Ahmed El-Kishky, Daniel Selsam, Francis Song, Giambattista Parascandolo, Hongyu Ren, Hunter Lightman, Hyung Won Chung, Ilge Akkaya, Ilya Sutskever, Jason Wei, Jonathan Gordon, Karl Cobbe, Kevin Yu, Lukas Kondraciuk, Max Schwarzer, Mostafa Rohaninejad, Noam Brown, Shengjia Zhao, Trapit Bansal, Vineet Kosaraju, Wenda Zhou, and OpenAI. Learning to Reason with LLMs, 2024. URL https://openai.com/index/learning-to-reason-with-llms/.

Epoch AI. Data on ai models, 07 2025. URL https://epoch.ai/data/ai-models. Accessed: 2025-09-24.

Jonathan St B. T. Evans. *Bias in Human Reasoning: Causes and Consequences*. Essays in Cognitive Psychology. Laurence Erlbaum Associates, 1994. ISBN 978-0-86377-156-9.

Leo Gao, Jonathan Tow, Baber Abbasi, Stella Biderman, Sid Black, Anthony DiPofi, Charles Foster, Laurence Golding, Jeffrey Hsu, Alain Le Noac'h, Haonan Li, Kyle McDonell, Niklas Muennighoff, Chris Ociepa, Jason Phang, Laria Reynolds, Hailey Schoelkopf, Aviya Skowron, Lintang Sutawika, Eric Tang, Anish Thite, Ben Wang, Kevin Wang, and Andy Zou. The language model evaluation harness, 07 2024. URL https://zenodo.org/records/12608602.

Vittorio Girotto, Andrea Mazzocco, and Alessandro Tasso. The effect of premise order in conditional reasoning: A test of the mental model theory. *Cognition*, 63(1):1–28, 1997. doi: 10.1016/S0010-0277(96)00792-5. URL https://pubmed.ncbi.nlm.nih.gov/9187063/.

Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021.

Manish Gupta and Puneet Agrawal. Compression of deep learning models for text: A survey. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16(4):1–55, 2022.

Abigail Z. Jacobs. Measurement as governance in and for responsible AI, 2021. URL http://arxiv.org/abs/2109.05658.

Philip Nicholas Johnson-Laird and Ruth M. J. Byrne. *Deduction*. Psychology Press, 1991.

P.N. Johnson-Laird. Mental Models, Sentential Reasoning, and Illusory Inferences. In *Advances in Psychology*, volume 138, pp. 27–51. Elsevier, 2006. ISBN 978-0-444-52079-1. doi: 10.1016/S0166-4115(06)80026-9. URL https://linkinghub.elsevier.com/retrieve/pii/S0166411506800269.

Daniel Kahneman and Amos Tversky. The Psychology of Preferences. *Scientific American*, 246(1):160–173, 1982. ISSN 0036-8733. doi: 10.1038/scientificamerican0182-160. URL https://www.scientificamerican.com/article/the-psychology-of-preferences.

Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy*, 98(6):1325–1348, 1990. ISSN 0022-3808, 1537-534X. doi: 10.1086/261737. URL https://www.journals.uchicago.edu/doi/10.1086/261737.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large Language Models are Zero-Shot Reasoners, 2022. URL https://arxiv.org/abs/2205.11916.

Philipp Koralus. *Reason and Inquiry: The Erotetic Theory*. Oxford University Press, 1 edition, 2022. ISBN 978-0-19-186254-0. URL https://doi.org/10.1093/oso/9780198823766.001.0001.

Philipp Koralus and Salvador Mascarenhas. The Erotetic Theory of Reasoning: Bridges Between Formal Semantics and the Psychology of Deductive Inference. *Philosophical Perspectives*, 27 (1):312–365, 2013. ISSN 1520-8583, 1520-8583. doi: 10.1111/phpe.12029. URL https://onlinelibrary.wiley.com/doi/10.1111/phpe.12029.

Philipp Koralus and Vincent Wang-Maścianica. Humans in Humans Out: On GPT Converging Toward Common Sense in both Success and Failure, 2023. URL http://arxiv.org/abs/2303.17276.

Philipp Koralus, Sean Moss, and Mark Todd. PyETR, 2025. URL https://github.com/Oxford-HAI-Lab/PyETR.

Andrew K Lampinen, Ishita Dasgupta, Stephanie C Y Chan, Hannah R Sheahan, Antonia Creswell, Dharshan Kumaran, James L McClelland, and Felix Hill. Language models, like humans, show content effects on reasoning tasks. *PNAS Nexus*, 3(7):pgae233, 2024. ISSN 2752-6542. doi: 10.1093/pnasnexus/pgae233. URL https://doi.org/10.1093/pnasnexus/pgae233.

Changmao Li and Jeffrey Flanigan. Task contamination: Language models may not be few-shot anymore. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(16):18471–18480, 2024. doi: 10.1609/aaai.v38i16.29808. URL https://ojs.aaai.org/index.php/AAAI/article/view/29808.

Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Alexander Cosgrove, Christopher D Manning, Christopher Re, Diana Acosta-Navas, Drew Arad Hudson, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue WANG, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yuksekgonul, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri S. Chatterji, Omar Khattab, Peter Henderson, Qian Huang, Ryan Andrew Chi, Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chaudhary, William Wang, Xuechen Li, Yifan Mai, Yuhui Zhang, and Yuta Koreeda. Holistic evaluation of language models. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=iO4LZibEqW. Featured Certification, Expert Certification.

Salvador Mascarenhas and Philipp Koralus. Illusory inferences: Disjunctions, indefinites, and the erotetic theory of reasoning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2015. URL https://escholarship.org/uc/item/1ft528sw.

Salvador Mascarenhas and Philipp Koralus. Illusory inferences with quantifiers. *Thinking and Reasoning*, 23(1):33–48, 2017a. doi: 10.1080/13546783.2016.1167125.

Salvador Mascarenhas and Philipp Koralus. Illusory inferences with quantifiers. *Thinking & Reasoning*, 23(1):33–48, 2017b. ISSN 1354-6783, 1464-0708. doi: 10.1080/13546783.2016.1167125. URL https://www.tandfonline.com/doi/full/10.1080/13546783.2016.1167125.

OpenAI. Introducing ChatGPT, 2022. URL https://openai.com/index/chatgpt/.

OpenAI. GPT-4 Technical Report, 2024. URL http://arxiv.org/abs/2303.08774.

OpenRouter. Openrouter. https://openrouter.ai, 2025. Accessed: 2025-05-13.

Yangjun Ruan, Chris J. Maddison, and Tatsunori Hashimoto. Observational scaling laws and the predictability of langauge model performance. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=On5WIN7xyD.

Abulhair Saparov and He He. Language models are greedy reasoners: A systematic formal analysis of chain-of-thought. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=qFVVBzXxR2V.

Freda Shi, Xinyun Chen, Kanishka Misra, Nathan Scales, David Dohan, Ed H Chi, Nathanael Schärli, and Denny Zhou. Large language models can be easily distracted by irrelevant context. In *International Conference on Machine Learning*, pp. 31210–31227. PMLR, 2023.

Aaditya K. Singh, Muhammed Yusuf Kocyigit, Andrew Poulton, David Esiobu, Maria Lomeli, Gergely Szilvasy, and Dieuwke Hupkes. Evaluation data contamination in LLMs: How do we measure it and (when) does it matter?, 2024. URL https://arxiv.org/abs/2411.03923.

Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974. ISSN 00368075, 10959203. URL http://www.jstor.org/stable/1738360.

Clare R. Walsh and P. N. Johnson-Laird. Co-reference and reasoning. *Memory & Cognition*, 32 (1):96–106, 2004. ISSN 0090-502X, 1532-5946. doi: 10.3758/BF03195823. URL http://link.springer.com/10.3758/BF03195823.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, 2023. URL http://arxiv.org/abs/2201.11903.

Fengli Xu, Qianyue Hao, Zefang Zong, Jingwei Wang, Yunke Zhang, Jingyi Wang, Xiaochong Lan, Jiahui Gong, Tianjian Ouyang, Fanjin Meng, Chenyang Shao, Yuwei Yan, Qinglong Yang, Yiwen Song, Sijian Ren, Xinyuan Hu, Yu Li, Jie Feng, Chen Gao, and Yong Li. Towards large reasoning models: A survey of reinforced reasoning with large language models, 2025. URL https://arxiv.org/abs/2501.09686.

Xiaohan Xu, Ming Li, Chongyang Tao, Tao Shen, Reynold Cheng, Jinyang Li, Can Xu, Dacheng Tao, and Tianyi Zhou. A survey on knowledge distillation of large language models, 2024. URL https://arxiv.org/abs/2402.13116.

Kun Zhou, Yutao Zhu, Zhipeng Chen, Wentong Chen, Wayne Xin Zhao, Xu Chen, Yankai Lin, Ji-Rong Wen, and Jiawei Han. Don't Make Your LLM an Evaluation Benchmark Cheater, 2023. URL http://arxiv.org/abs/2311.01964.

A A FORMAL PRESENTATION OF THE EROTETIC THEORY

ETR is formulated in Koralus (2022) in a set-theoretic language, and we reproduce the full definitions here, with the explicit permission of the authors. We remark that Definitions 1 and 11 contain the parameters of the theory, i.e. the basic atomic propositions, the constants and function symbols, and a set logical and extralogical axioms. The central definition is that of a *View*, Definition 9. Following that are definitions of the basic operations of ETR, in particular Update (Definition 23) and Query (Definition 30) whose use in PyETR were demonstrated in Section 2.

Definition 1 (**Basic objects**).
Let $\mathbf{A}, \mathbf{F}, \{?\}$, $\mathbf{P}^\top$, $\mathbf{P}^\perp$ *be pairwise disjoint countable sets.*

(**F** *arity*) $\alpha : \mathbf{F} \to \mathbb{N}$. *Write* $f \in \mathbf{F}$ *as* f^k *when* $\alpha(f) = k$.
(Constants) $\exists w_{\in \mathbf{F}}\ \alpha(w) = 0$.
(**P** *polarity*) *Let* $N : \mathbf{P}^\top \to \mathbf{P}^\perp$ *be a bijection. Write* $\bar{P}$ *for* $N(P_{\in \mathbf{P}^\top})$.
Let $\mathbf{P} = \mathbf{P}^\top \cup \mathbf{P}^\perp$.
(**P** *arity*) $\alpha' : \mathbf{P} \to \mathbb{N}$. *Write* P *as* P^k *for* $\alpha'(P) = k$. $\alpha'(P) = \alpha'(\bar{P})$.
(Identity) $=^2 \in \mathbf{P}$.

Definition 2 (**Terms T**).
$\mathbf{A} \subseteq \mathbf{T}$. *If* $f^0 \in \mathbf{F}$, *then* $\langle f^0 \rangle \in \mathbf{T}$. *If* $f^k \in \mathbf{F} \wedge \{t_1 \dots t_k\} \subseteq \mathbf{T}$, *then* $\langle f^k, \langle t_1 \dots t_k \rangle \rangle \in \mathbf{T}$.

Definition 3 (**Atoms** $\mathcal{A}$).
If $P^k \in \mathbf{P}$ *and* $\vec{\tau} = \langle t_1 \dots t_k \rangle$ *for* $t_i \in \mathbf{T}$, *then* $\langle P^k, \vec{\tau} \rangle \in \mathcal{A}$ *(abbrv.* $P\vec{\tau}$*).* $|\vec{\tau}| = \{t_1 \dots t_k\}$.

Definition 4 (**States** $\mathbb{S}$).
$\mathbb{S} = \mathcal{P}(\mathcal{A})$. *Abbreviate* '{ }' *as* '0'.

Definition 5 (**Dependency relations** $\mathcal{R}$).
Let $\Gamma \in \mathcal{P}(\mathbb{S})$. *Let* $U, E \subseteq \mathbf{A}(\Gamma)$ *and* $D \subseteq E \times U = \{\langle e, u\rangle : e \in E \wedge u \in U\}$. *Define* $\langle U, E, D\rangle \in \mathcal{R}_\Gamma$ *iff*

(Bipartite) $U \cap E = \emptyset \wedge D \subseteq E \times U$, *and*
(Matryoshka) *For all* $u, u' \in U$ $\{e \in E : \langle e, u\rangle \in D\} \subseteq \{e \in E : \langle e, u'\rangle \in D\}$ *or* $\{e \in E : \langle e, u'\rangle \in D\} \subseteq \{e \in E : \langle e, u\rangle \in D\}$.

For $R \in \mathcal{R}$ *write* $R = \langle U_R, E_R, D_R\rangle$. *For* $\langle \emptyset, \emptyset, \emptyset\rangle$ *write* $0_\mathcal{R}$.

Definition 6 (**Open terms** $\mathbf{T_1}$**and atoms** $\mathcal{A}_1$).
$? \in \mathbf{T}_1$ *and* $\langle f^k, \langle t_1, \ldots, t_k\rangle\rangle \in \mathbf{T}_1$ *whenever* $f^k \in \mathbf{F}$ *and for some* $1 \leq i \leq k$, $t_i \in \mathbf{T}_1$ *and* $t_j \in \mathbf{T}$ *for* $j \neq i$. $\langle P^k, \langle t_1, \ldots, t_k\rangle\rangle \in \mathcal{A}_1$, *whenever* $P^k \in \mathbf{P}$ *and for some* $1 \leq i \leq k$, $t_j \in \mathbf{T}$ *for all* $j \neq i$ *and* $t_i \in \mathbf{T}_1$.

Definition 7 (**Issue structures** $\mathbb{I}$).
$I \in \mathbb{I}_\Gamma$ *iff* I *consists of pairs* $\langle t, x\rangle$, *s.t.* $x \in \mathcal{A}_1$ *and* $x[t/?] \in \mathcal{A}(\Gamma)$.

Definition 8 (**Issue matches** M_{IJ}).
For $I, J \in \mathbb{I}_\Gamma$, *define,*

$$M_{IJ} = \{\langle t_1, t_2\rangle : \exists x(\langle t_1, x\rangle \in I \wedge (\langle t_2, x\rangle \in J \vee \langle t_2, \bar{x}\rangle \in J))\}$$

where, for $x = \langle P^k, \langle t_1, \ldots, t_k\rangle\rangle$, $\bar{x} = \langle \bar{P}^k, \langle t_1, \ldots, t_k\rangle\rangle$.

Definition 9 (**Views** $\mathbb{V}$).
For Γ, Θ *finite subsets of* $\mathbb{S}$, $R \in \mathcal{R}_{\Gamma\cup\Theta}$, $I \in \mathbb{I}_{\Gamma\cup\Theta}$,
$\langle \Gamma, \Theta, R, I\rangle \in \mathbb{V}$ *(abbreviated* $\Gamma^\Theta_{RI} \in \mathbb{V}$*).*
Write $\top$ *for* $\{0\}^{\{0\}}_{0_\mathcal{R}\emptyset}$ *and* $\bot$ *for* $\emptyset^{\{0\}}_{0_\mathcal{R}\emptyset}$.

Definition 10 (**Commitments** $\mathbb{C}$).
For $C \subseteq \mathbb{V}, G \in \mathbb{V}$, $\langle C, G\rangle \in \mathbb{C}$.

Definition 11 (**Primitive absurd states** $\mathbb{K}$).
Let $\mathbb{K}_{\subseteq\mathbb{S}}$ *contain,* $\forall t, t'_{\in\mathbf{T}} \forall p_{\in\mathcal{A}} \forall x_{\in\mathcal{A}_1}$, *at least:*

(LNC) $\{p, \bar{p}\}$,
(Aristotle) $\{\neq tt\}$,
(Leibniz) $\{=tt', x[t/?], \bar{x}[t'/?]\}$.

Definition 12 ($\mathcal{R}$ **Restriction**).
$[R]_X = \langle U_R \cap X, E_R \cap X, D_R \cap ((E_R \cap X) \times (U_R \cap X))\rangle$.
$[R]_\Gamma = [R]_{\mathbf{A}(\Gamma)}$.
Given $\langle \Gamma, \Theta, R, I\rangle$, *we allow ourselves to write* $\langle \Gamma, \Theta, [R], I\rangle$ *for* $\langle \Gamma, \Theta, [R]_{\Gamma\cup\Theta}, I\rangle$.

Definition 13 ($\mathbb{I}$ **Restriction**).
Within a quadruple $\langle \Gamma, \Theta, R, [I]\rangle$, *let*
$[I] = \{\langle t, x\rangle : \langle t, x\rangle \in I \wedge x[t/?] \in \mathcal{A}(\Gamma \cup \Theta)\}$

Definition 14 ($\mathcal{R}$ **Algebra**).
Let $R \rtimes S = \langle U_R \cup U_S, E_R \cup E_S, D_R \cup D_S \cup E_S \times U_R\rangle$ *and let*
$0_\mathcal{R} \bowtie 0_\mathcal{R} = 0_\mathcal{R}$.
Let

$$R \bowtie S = \langle U_0, E_0, \emptyset\rangle \rtimes ([R]_{\mathbf{A}(R)-(E_0\cup U_0)} \bowtie [S]_{\mathbf{A}(S)-(E_0\cup U_0)})$$

where $E_0 = \{e_{\in E_R\cup E_S} : \forall u.\langle e, u\rangle \notin D_R \cup D_S\}$ *and*
$U_0 = \{u_{\in U_R\cup U_S} : \forall e \notin E_0.\langle e, u\rangle \notin (E_R \times U_R - D_R) \cup (E_S \times U_S - D_S)\}$.

Definition 15 (**Product**).
$\Gamma^\Theta \otimes \Delta^\Psi = \big(\{\gamma_{\in\Gamma} \cup \delta_{\in\Delta} : \exists\psi_{\in\Psi}(\psi \subseteq \gamma)\} \cup \{\gamma_{\in\Gamma} : \neg\exists\psi_{\in\Psi}(\psi \subseteq \gamma)\}\big)^\Theta$
$(\Gamma \otimes \Delta)^{\{0\}} = \Gamma^{\{0\}} \otimes \Delta^{\{0\}}$
$\bigotimes_{i\in\mathbf{P}} \Delta_i^{\Psi_i} = \{0\}^{\{0\}} \otimes \Delta_1^{\Psi_1} \otimes \ldots \otimes \Delta_n^{\Psi_n}$
$\Gamma^\Theta_{RI} \otimes^T \Delta^\Psi_{SJ} = (\Gamma^\Theta \otimes \Delta^\Psi)_{[(T\bowtie R)\bowtie(T\bowtie S)][I\cup J]}$
$\Gamma^\Theta_{RI} \otimes \Delta^\Psi_{SJ} = \Gamma^\Theta_{RI} \otimes^{0_\mathcal{R}} \Delta^\Psi_{SJ}$
$\bigotimes^T_{i\in\mathbf{P}} {\Delta_i}^{\Psi_i}_{S_iJ_i} = \top \otimes^T {\Delta_1}^{\Psi_1}_{S_1J_1} \otimes^T \ldots \otimes^T {\Delta_n}^{\Psi_n}_{S_nJ_n}$

Definition 16 (Sum).
$\Gamma^{\Theta}_{RI} \oplus^{T} \Delta^{\Theta}_{SJ} = (\Gamma \cup \Delta)^{\Theta}_{[(T \bowtie R) \bowtie (T \bowtie S)][I \cup J]}$
$\Gamma^{\Theta}_{RI} \oplus \Delta^{\Theta}_{SJ} = \Gamma^{\Theta}_{RI} \oplus^{0_{\mathcal{R}}} \Delta^{\Theta}_{SJ}$
$\bigoplus^{T}_{i \in \mathbf{P}} \Delta^{\Psi_i}_{S_i J_i} = \emptyset^{\Psi}_{0_{\mathcal{R}}\emptyset} \oplus^{T} {\Delta_1}^{\Psi}_{S_1 J_1} \oplus^{T} \ldots \oplus^{T} {\Delta_n}^{\Psi}_{S_n J_n}$
Definition 17 ($\mathcal{A}$ Answer potential).
$\Gamma[\Delta]^{\mathcal{A}P} = |\mathcal{A}(\Gamma) \cap \mathcal{A}(\Delta)|$
Definition 18 (Answer).

$$\Gamma^{\Theta}_{RI}[\Delta^{\{0\}}_{SJ}]^{A} = (\underset{\gamma \in \Gamma}{\arg\max}\ \Delta[\{\{p\} : p \in \gamma\}]^{\mathcal{A}P})^{\Theta}_{[R][I]}$$

Definition 19 (Negation).
$[\Gamma^{\Theta}_{RI}]^{N} = (\Theta^{\{0\}} \otimes ([\Gamma]^{N})^{\{0\}})_{[R]^{N}[I]^{N}}$
$[\Gamma]^{N} = \bigotimes_{\gamma \in \Gamma} \hat{\gamma}^{\{0\}}$
$\hat{\gamma} = \big\{\{\bar{p}\} : p_{\in \gamma}\big\}$
$\bar{p} = \bar{F}\vec{\tau}$ *if* $p = F\vec{\tau}$; $\bar{p} = F\vec{\tau}$ *if* $p = \bar{F}\vec{\tau}$
$[R]^{N} = \langle E_R, U_R, \{\langle a, b\rangle \in U_R \times E_R : \langle b, a\rangle \notin D_R\}\rangle$
$[I]^{N} = I \cup \{\langle t, \bar{x}\rangle : \langle t, x\rangle \in I\}$
Definition 20 (Novelty)**.** $a \in \mathbf{A}$ *is* novel *during an inference step* $\langle C, G\rangle[D]^{O}$ *if* $a \notin \mathbf{A}(G) \cup \mathbf{A}(D)$ *and* a *has not yet appeared in the computation of the conclusion of the step. A function* $\nu : X \to \mathbf{A}$*, where* $X \subseteq \mathbf{A}$ *is finite, is novel during an inference step if each* $\nu(x)$ *is novel and* ν *is injective, and then* $[\nu]_X$ *denotes the simultaneous substitution* $[\nu(a_1)/a_1, \ldots, \nu(a_k)/a_k]$*, where* $X = \{a_1, \ldots, a_k\}$*. Let* $(-)^{\mathrm{nov}(X)}$ *stand for* $(-)[\nu]_X$ *where* $\nu : X \to \mathbf{A}$ *is novel.*
Definition 21 (Substitution)**.** *Let* $Z(T, a) = \{u \in U_T : u \lhd_T a\} \cup \{e \in E_T : e \lesssim_T a\} - \{a\} = \{a_1, \ldots, a_k\}$*. For* $\mathbf{A}(\Gamma) \cap \mathbf{A}(\Theta) = \emptyset$*,*

$$\mathrm{Sub}^{T}_{\langle t, a\rangle}(\Gamma^{\Theta}_{I}) = \langle \Gamma[\nu_1]_{Z(T,a)}[t/a], \Theta, [T \rtimes ([T]_{Z(T,a)}[\nu_1]_{Z(T,a)})], I[\nu_1]_{Z(T,a)}[t/a]\rangle,$$

where $\nu_1 : Z(T, a) \to \mathbf{A}$ *is novel.*
Definition 22 (Merge).
For either $\mathbf{A}(R) \cap \mathbf{A}(S) = \emptyset$ *or* $[R]_{\Delta \cup \Psi} = S$*,*
$\Gamma^{\Theta}_{RI}[\Delta^{\Psi}_{SJ}]^{M} = \bigoplus^{R \bowtie S}_{\gamma \in \Gamma} \Big(\{\gamma\}^{\Theta}_{RI} \otimes^{R \bowtie S} \Delta^{\Psi}_{SJ} \otimes^{R \bowtie S} \bigotimes^{R \bowtie S}_{\langle t, u\rangle \in M'_{IJ}(\gamma)} \mathrm{Sub}^{R \bowtie S}_{\langle t, u\rangle}(\Delta^{\{0\}}_{J})\Big),$
where $M'_{IJ}(\gamma) = \{\langle t, u\rangle \in M_{IJ} : u \in U_S \wedge \exists \psi_{\in \Psi}(\psi[t/u] \subseteq \gamma \wedge \psi \not\subseteq \gamma)\}$
Definition 23 (Update).
For $D \in C$*, but with all arbitrary objects novelised,*
$\langle C, \Gamma^{\Theta}_{RI}\rangle[D]^{\circlearrowright} = \langle C, \Gamma^{\Theta}_{RI}[D]^{U}[D]^{E}[D]^{A}[D]^{M}\rangle$
Definition 24 (Universal product).
For $\mathbf{A}(\Gamma) \cap \mathbf{A}(\Theta) = \emptyset$ *and either* $\mathbf{A}(R) \cap \mathbf{A}(S) = \emptyset$ *or* $[R]_{\Delta} = S$*,*
$\Gamma^{\Theta}_{RI}[\Delta^{\{0\}}_{SJ}]^{U} = \{0\}^{\Theta}_{RI} \otimes^{R \bowtie S} \bigotimes^{R \bowtie S}_{\langle u, t\rangle \in M'_{IJ}} \mathrm{Sub}^{R \bowtie S}_{\langle t, u\rangle}(\Gamma^{\{0\}}_{I})$*, where*
$M'_{IJ} = \{\langle u, t\rangle : \langle u, t\rangle \in M_{IJ} \wedge u \in U_R - \mathbf{A}(\Theta)\} \neq \emptyset$
Definition 25 (Reorient).
For $\Delta^{\Psi}_{SJ} \in (C \cup \{\Gamma^{\Theta}_{RI}\})$ *and* $J' \in \mathbb{I}_{\Delta \cup \Psi}$*,*
$\langle C, \Gamma^{\Theta}_{RI}\rangle[\Delta^{\Psi}_{SJ'}]^{R} = \langle C, \Delta^{\Psi}_{SJ'}\rangle$
Definition 26 (Existential sum).
Let the following conditions be met:

(1) $\mathbf{A}(\Gamma) \cap \mathbf{A}(\Theta) = \emptyset$ *and either* $\mathbf{A}(R) \cap \mathbf{A}(S) = \emptyset$ *or* $[R]_{\Delta} = S$*.*

(2) $M'_{IJ} = \{\langle e, t\rangle : \langle e, t\rangle \in M_{IJ} \wedge e \in E_R - \mathbf{A}(\Theta \cup \Delta) \wedge \neg\exists x(\langle e, x\rangle \in D_R)\} \neq \emptyset$*.*

Then,

$$\Gamma^{\Theta}_{RI}[\Delta^{\{0\}}_{SJ}]^{E} = \Gamma^{\Theta}_{RI} \oplus^{R \bowtie S} \bigoplus^{R \bowtie S}_{\langle e, t\rangle \in M'_{IJ}} \mathrm{Sub}^{R \bowtie S}_{\langle t, e\rangle}\Big(\big(\bigcup_{\gamma \in \{\gamma_{\in \Gamma} : e \in \mathbf{A}(\gamma)\}} (\{\gamma\} \cup \big\{\{x_{\in \gamma} : e \notin \mathbf{A}(x)\} \cup \delta : \delta \in \bigotimes_{x \in B(\gamma, I, e)} \{\{x\}, \{\bar{x}\}\} \wedge \delta \not\subseteq \gamma\}\big)\big)^{\Theta}_{I}\Big),$$

where for $e \in E_R$ and $\gamma \in \Gamma$, let $B(\gamma, I, e) = \{x[e/?] \in \gamma : \langle e, x\rangle \in I\}$.

Definition 27 (**Division**).
If $\forall\delta_{\in\Delta}\exists\psi_{\in\Psi}\exists\gamma_{\in\Gamma}(\delta \subseteq \gamma \wedge \psi \subseteq \gamma)$, then
$\Gamma^{\Theta}_{RI} \oslash \Delta^{\Psi}_{SJ} = \{\gamma \oslash_{\Gamma} \Delta^{\Psi} : \gamma \in \Gamma\}^{\Theta}_{[R][I]}$
$\gamma \oslash_{\Gamma} \Delta^{\Psi} = \gamma - \iota\delta(\delta \in \Delta \wedge \delta \subseteq \gamma \wedge \exists\psi_{\in\Psi}(\psi \subseteq \gamma))$

Definition 28 (**Factor**).
For $\Delta^{\Psi}_{SJ} \neq \bot$,
$\Gamma^{\Theta}_{RI}[\Delta^{\Psi}_{SJ}]^F = \{\gamma[\Delta^{\Psi}]^F : \gamma \in \Gamma\}^{\Theta}_{[R][I]}$ *where*
$\gamma[\Delta^{\Psi}]^F = (\gamma \oslash_{\Gamma} \Delta^{\Psi}) \cap \bigcap\{\gamma \oslash_{\Gamma} (\Delta^{\Psi}[t/a]) : \langle t, a\rangle \in M_{IJ} \wedge a \in U_S\}$*, and where we let* $\gamma' \cap \bigcap \emptyset = \gamma'$.
For $\Delta^{\Psi}_{SJ} = \bot$,
$\Gamma^{\Theta}_{RI}[\bot]^F = \left\{\gamma \in \Gamma : \neg\exists\delta \in \mathbb{K}.\delta \subseteq \gamma\right\}^{\Theta}_{[R][I]}$.

Definition 29 (Matryoshka level ordering).
For existentials, we have $e \lesssim_R e'$ iff $\forall\langle e', u\rangle \in D_R(\langle e, u\rangle \in D_R)$. For both existentials and universals, let the relation $\lhd_R$ be the transitive closure of the relation made up by the ordered pairs in $D_R \cup (E_R \times U_R - D_R)^{\mathrm{op}}$, where op reverses the order of pairs in a set of pairs.

Definition 30 (**Query**).
For $U_S \subseteq U_R$ and $M'_{IJ} = \{\langle t, e\rangle \in M_{IJ} : e \in E_S \backslash E_R\}$, we define
$\Gamma^{\Theta}_{RI}[\Delta^{\Psi}_{SJ}]^Q =$

$$(\{0 : \neg\exists\delta_{\in\Delta}\exists\gamma.\Phi(\gamma, \delta)\} \cup \{\delta \in \Delta : \exists\gamma_{\in\Gamma}.\Phi(\gamma, \delta)\})^{\Theta}_{[R\bowtie\langle U_R, E_S\backslash E_R, D_{S'}\rangle][I\cup J]}$$

where $\Phi(\gamma, \delta) \leftrightarrow \exists\psi_{\in\Psi}\exists n_{\geq 0}\exists\langle t_1, e_1\rangle, \ldots, \langle t_n, e_n\rangle_{\in M'_{IJ}}\forall i, j$
$$(\psi \cup \delta[t_1/e_1, \ldots, t_n/e_n] \subseteq \gamma \wedge e_i = e_j \rightarrow i = j);$$
and $D_{S'} = D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5 \cup D_6$ is constructed by taking: original dependencies from the internal argument
$$D_1 = [D_S]_{\mathbf{A}/E_R},$$
extra dependencies for multiple terms substituted with same e
$$D_2 = \{\langle e_m, u\rangle : \exists m\exists m' \in M'_{IJ}(e_m = e_{m'} \wedge t_m \neq t_{m'}) \wedge u \in U_R\},$$
dependencies resulting from complex terms being substituted
$$D_3 = \{\langle e_m, u\rangle : \langle t_m, e_m\rangle \in M'_{IJ} \wedge u \in U_R(t_m)\},$$
$$D_4 = \{\langle e_m, u\rangle : \langle t_m, e_m\rangle \in M'_{IJ} \wedge e \in E_R(t_m) \wedge \langle e, u\rangle \in R\},$$
$$D_5 = \{\langle e_m, u\rangle : \langle t_m, e_m\rangle \in M'_{IJ} \wedge u \in U_R \wedge \forall u'_{\in U_R(D_3\cup D_4)}(u' \lhd_R u)\},$$
additional dependencies necessary to preserve the dependency order of existentials in S
$$D_6 = \{\langle e, u\rangle : e, e' \in E_S - E_R \wedge e \lesssim_S e' \wedge$$
$$(\forall m, m' \in M'_{IJ}(e_m = e' \wedge e_{m'} = e') \rightarrow t_m = t_{m'}) \wedge \langle e', u\rangle \in D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5\}.$$

Definition 31 (**Wh-Query**). *Provided that $U_S \subseteq U_R$,*
$\Gamma^{\Theta}_{RI}[\Delta^{\Psi}_{SJ}]^W = \Big(\left\{0 : \exists\gamma \in \Gamma.\neg\exists\delta_{\in\Delta}.\Phi(\gamma, \delta)\right\} \cup \left\{\delta \in \Delta : \exists\gamma_{\in\Gamma}.\Psi(\gamma, \xi)\right\}\Big)^{\Theta}_{[R][I]}$*, where $\Psi(\gamma, \xi)$ iff*

$$\exists\psi_{\in\Psi}\exists n_{\geq 0}\exists\langle t_1, e_1\rangle, \ldots, \langle t_n, e_n\rangle_{\in M'_{IJ}}\exists\delta_{\in\Delta}.$$
$$\big(\xi \cup \psi \subseteq \gamma \wedge \xi = \delta[t_1/e_1, \ldots, t_n/e_n]$$
$$\wedge \forall i, j.(e_i = e_j \rightarrow i = j)\big).$$

Definition 32 (**Inquire**).
(O) If $\mathbf{A}(\Gamma \cup \Theta) \cap \mathbf{A}(\Delta \cup \Psi) = \emptyset \wedge \mathbf{A}(\Delta) \cap \mathbf{A}(\Psi) = \emptyset$, then
$\Gamma^{\Theta}_{RI}[\Delta^{\Psi}_{SJ}]^I = \Gamma^{\Theta}_{RI} \otimes \Big(\Delta^{\Psi}_{SJ} \oplus \big(\{0\}^{\Psi}_{[S][J]} \otimes Nov([\Delta^{\{0\}}]^N_{[[S]^N][J]^N})\big)\Big)[\bot]^F$,
where $Nov()$ uniformly replaces all arbitrary objects in its scope by novel ones.

(I) If $\mathbf{A}(\Delta \cup \Psi) \subseteq \mathbf{A}(\Gamma \cup \Theta)$, then
$\Gamma^{\Theta}_{RI}[\Delta^{\Psi}_{SJ}]^I = \Gamma^{\Theta}_{RI} \otimes \big(\Delta^{\Psi}_{0_{\mathcal{R}}J} \oplus ([\Delta]^N_{0_{\mathcal{R}}[J]^N})^{\Psi}\big)[\bot]^F$.

Definition 33 (**Suppose**).
(O) If $\mathbf{A}(\Gamma \cup \Theta) \cap \mathbf{A}(\Delta \cup \Psi) = \emptyset$*, then*
$\Gamma^{\Theta}_{RI}[\Delta^{\Psi}_{SJ}]^{S} = \Gamma^{\Theta'}_{[R \bowtie R'][I \cup I']}[\Delta^{\Psi}_{SJ}]^{U}[\Delta^{\Psi}_{SJ}]^{E}[\Delta^{\Psi}_{SJ}]^{A}[\Delta^{\Psi}_{SJ}]^{M}$*, where*
$\Theta'^{\{0\}}_{R'I'} = \Theta^{\{0\}}_{RI} \otimes Nov(\Delta^{\Psi}_{[S]^{N}J}[\,]^{D})$,
where $Nov()$ *uniformly replaces all arbitrary objects in its scope by novel ones.*

(I) If $\mathbf{A}(\Delta) \subseteq \mathbf{A}(\Gamma \cup \Theta)$ *and* $[R]_{\Delta} = S$*,*
then $\Gamma^{\Theta}_{RI}[\Delta^{\{0\}}_{SJ}]^{S} = \Gamma^{\Theta \otimes \Delta}_{RI}[\Delta^{\{0\}}_{SJ}]^{U}[\Delta^{\{0\}}_{SJ}]^{E}[\Delta^{\{0\}}_{SJ}]^{A}[\Delta^{\{0\}}_{SJ}]^{M}$

Definition 34 (**Depose**).
$\Gamma^{\Theta}_{RI}[\top]^{D} = \langle \Gamma^{\{0\}} \oplus ([\Theta]^{N})^{\{0\}}, R, [[I]^{N} \cup I] \rangle$.

Definition 35 (**Commit**).
$\langle C, \Gamma^{\Theta}_{RI} \rangle [\top]^{C} = \langle C \cup \{\Gamma^{\Theta}_{RI}\}, \Gamma^{\Theta}_{RI} \rangle$

Definition 36 (Inference). *An inference* $\mathcal{I}$ *is a finite sequence of inference steps defined as follows, relative to an inference state consisting of a commitment* $\langle C, G \rangle \in \mathbb{C}$*. If* $E \in \mathbb{V}$ *and* $O \in \{[D]^{\circlearrowleft}, [D]^{S}, [D]^{Q}, [D]^{F}, [D]^{R}, [D]^{I}, [D]^{D}, [D]^{E}, [D]^{U}, [D]^{C}\}$*, then* O *is an inference step applicable to* $\langle C, G \rangle$*, whose result is an inference state in* $\mathbb{C}$*. We write* $C \vdash_{\mathrm{ETR}} E$ *if and only if there is an erotetic theory inference* $\mathcal{I}$ *s.t.* $\langle C, \top \rangle \mathcal{I} = \langle C', E' \rangle$*, where* E' *is identical to* E *up to uniform replacement of* $\mathbf{A}$ *objects.*